DOCUMENT RESUME

ED 181 034 TM 009 909

AUTHOR Donlon, Thomas F.
TITLE Time for Review as an Aspect of Test Speededness.
PUB DATE 79
NOTE 12p.: Paper presented at the Annual Meeting of the New England Educational Research Organization (Provincetown, MA, April 29-May 1, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Entrance Examinations: Higher Education: Proctoring: Response Style (Tests): *Review (Reexamination): Senior High Schools: Statistical Analysis: Technical Reports: Testing: Testing Problems: *Time: *Timed Tests
IDENTIFIERS *Speededness (Tests)

ABSTRACT

Questions are raised about the time within a test administration period that may be available for review for those candidates who have time enough, or more than enough, to finish. The assumption is proposed that rate of work may be normally distributed. The rate of work for those who have finished a test cannot be determined from their answer sheets; but the distribution of the rates of work for those who do not finish can be computed and, if rate of work is normally distributed, the distribution for remaining candidates may be extrapolated. Graphic and numerical methods are proposed for estimating the rate of work distribution for the complete candidate group, based on data derived from the portion of the group who do not complete the text. Questions are raised concerning the effects of candidates having more than enough time to complete a test. Can this time be effectively used to improve test scores? How many candidates use this time for this purpose? What other effects may result from candidates having more than enough time to complete a test? (CTM)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED181034

# Time for Review as an Aspect of Test Speededness

Thomas F. Donlon

Educational Testing Service
Princeton, NJ. 08540

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Thomas F. Donlon

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM009 909

New England
Educational Research Organization
1979

## Time for Review as an Aspect of Test Speededness

The assumption of normally distributed test scores is a familiar one. Assumptions about the rate of work of people on tests, however, are not so common. In fact, the distribution of rates of work is seldom formally considered. Donlon and Reilly (1976) have considered the usefulness of an assumption of a normally distributed rate of work. They use the assumption to infer total-group speed-power correlations from the test results of those failing to complete.

The approach to speededness that builds on an assumed normal distribution of work rates, however, has a potentially valuable by-product. To date, most comment on test speededness has centered on the problem of failure to finish. In this context, the basic problem is the usefulness and the fairness of the resulting score distribution, which blends the attainments of finishers and non-finishers. The question asked is: Are distinctions among individuals which result from rate of work, as evidence by a failure to finish, valid and meaningful? Can these distinctions be combined with distinctions which are based not on rate of work but on power, on success on items? But a related question might be: Are distinctions among individuals which result from rate of work, as evidenced by the time remaining upon completion, valid and meaningful? Does this "waiting time" make a difference? These distinctions are made among individuals who complete the test, these for whom the test is usually considered a "power-experience". But do they influence the power measures?

If the assumption of normality is sustained, there is potentially a wide range of rates of work within the two groups defined by completion. These differences affect the entire test experience. In particular, they lead directly to differences among candidates in the time available for review.

The time available for review is the time remaining, under the official allotment for the test, subsequent to the candidate's finishing the test. Little is known about this variable as a factor in test performance. The emphasis in speed analysis on the failure to complete has largely obscured the question of time available for review. Furthermore, there has not been a direct indication of this time derivable through an item or test analysis in a manner analogous to the use of the Not Reached variable as a measure of failure to complete. The assumption of normality, however, makes it possible to estimate times for review, for the assumption makes it possible to estimate the times for completion.

This paper presents the general logic for the estimation, from item analysis data, of the time-for-review distribution, its methodology and equations, and some demonstrations of the method using sections of the College Board Scholastic Aptitude Test.

Outline of the Method:

Donlon and Reilly proposed that rates of work were normally distributed and the mean and variance of the total distribution could be estimated, given this assumption, by using the observed work rates of subjects who had to finish. Only subjects who fail to finish offer evidence for a defined work rate; if a subject completes the test, there is no way of knowing how long that subject waited until time was called. But if a subject fails to finish, he or she offers a specific estimate of work rate. Thus, if the test is a 50-item test, and the time limit is 30 minutes, those who finish the test have work rates of 50/30=1.67 items/minute or greater. But those who finish 40 items have work rates of 40/30=1.33 items/minute, etc. Those not finishing the test constitute a subgroup of the assumed normal distrubution. By calculating the mean and standard derivation of the

observed rates of work for non-finishers, it is possible to estimate the mean and standard derivation for the whole group, because under the assumption of a

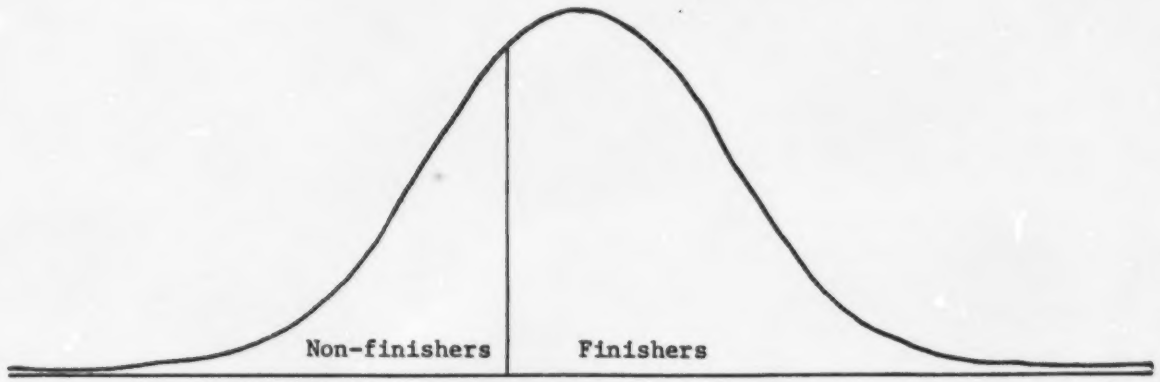


Figure 1. Rate of Work

normal distribution the observed values for non-finishers may be equated to moments of the truncated normal distribution. Appendix A offers the equations for this estimation.

It is also possible to estimate the work rates by graphic methods, plotting the data on normal probability paper, as indicated in Figure 2. In this method, the mean work rate is a function of the items completed or completable by 50% of the sample, as estimated by the point where the line crosses the 50% vertical. The standard deviation is based on the difference between the 50% vertical and the 84% vertical. In Figure 2, the 50% point is indicated by item 54; since the subjects were given 30 minutes, this is a rate of work of 1.8 items/minute. The 84% vertical is roughly equivalent to 42 items, a work rate of 42/30= 1.4 items per minute; thus, the standard deviation is .4 items/minute.

Time for review is a direct function of work rate, for time for review is the difference between the time allowed for the examination and the time required to complete it:

$$T_R = T_A - T_C$$

$T_R$ = time for review

$T_A$ = time allowed

$T_C$ = time required to complete

These time values may be conveniently expressed in a number of ways as items/minute, items/allowed time, etc. $T_C$ is determined by the equation.

$$T_C = \frac{n}{r} \quad \text{where}$$

n = the number of items in the test

r = the rate

The distribution of the time to review will not ordinarily be normal, given the assumption of a normally distributed rate of work, for the $T_C$ values are an inverse, rather than a linear function of the work rates. Consider the situation projected in Figure 2. The group confronted a 65-item test. On the basis of the estimates of a mean of 54 items completable and a standard deviation of 12 items, about 18% of the group would be expected to finish the test (about 21% actually did). But when did the earliest finishers complete the test? On the assumption of normality, about 1% of the sample have work rates 2.33 standard deviation units above the mean or greater. Such work rates in this instance would be 54 + (2.33) (12) = 82 items in 30 minutes, or 2.73 items/minute. Solving for $T_C$ yields 65/2.6 = 23.8 minutes. Thus, the earliest finishers on this test had about 6.2 minutes or more for review.

This general logic can be used to work out a distribution of time for review. Five minutes for review requires that the test be completed in 25

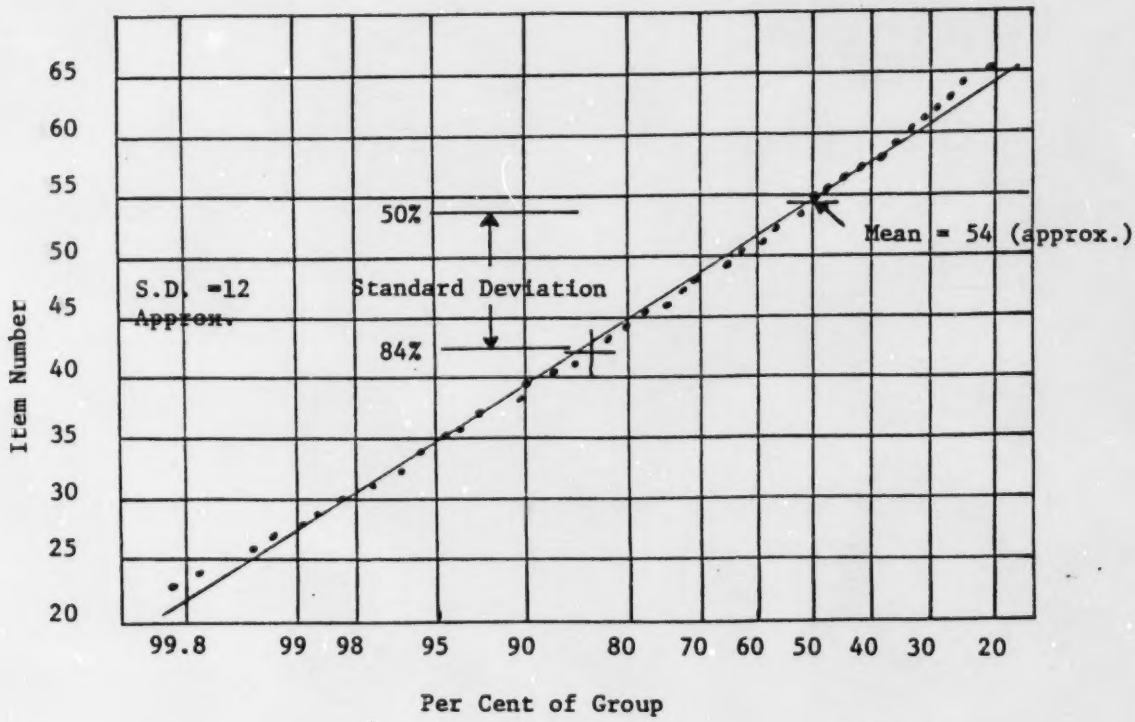


Figure 2. Estimation by graphic methods of the mean and standard deviation of items completable within the allotted time - RSA45 Sentence Completion

minutes. This is a rate of 65/25 = 2.6 items/minute, or, in terms of the "items completable" calculation, 78 items in the allowed time. This is 2.0 standard deviation units above the mean. About 2.3% of the group will show rates as great or greater than this. Four minutes for review requires completion in 26 minutes, a rate of 2.5 items/minute. About 4% of the group show rates as great or greater. These calculations can be continued, and a table developed, as in Table 1.

TABLE 1

Estimated time for review

$RSA45_{17}$ - SAT Sentence Completion Pretest

| Time for review | % in category |
|---|---|
| 6.0 min. or more | 1.2% |
| 5.0 - 5.99 | 1.1% |
| 4.0 - 4.99 | 1.7% |
| 3.0 - 3.99 | 2.0% |
| 2.0 - 2.99 | 4.0% |
| 1.0 - 1.99 | 4.0% |
| 0.0 - 0.99 | 4.0% |

These data give an indication of the distribution of time remaining for review. In this case, since most of the group did not finish the test, the percentages enjoying significant times for review are not great. Figure 3 presents data on a test that was essentially completed by the group working on it. When the methods are applied to the data of Figure 3, the distribution presented in Table 2 is determined.

In Table 2, the percentages of the group enjoying significant time for review are much greater. About 19% of the group may be estimated to have 5 minutes or more for review. The earliest finishers will have 8 minutes or more to go over the test.

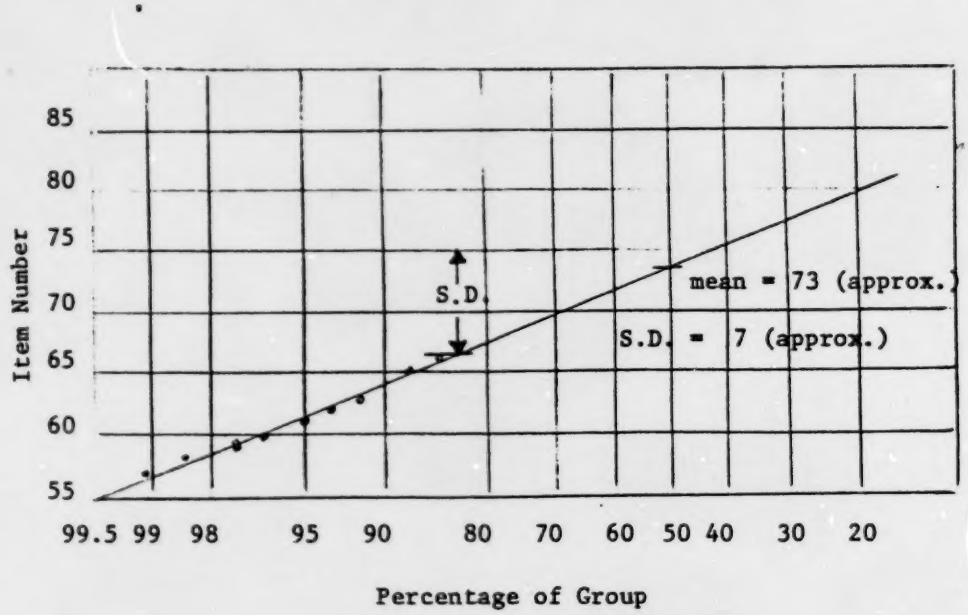


Figure 3. Estimation by graphic methods of the mean and standard deviation of items completable within the allotted time - RSA $45_{24}$ - verbal pretest.

TABLE 2

ESTIMATED TIME FOR REVIEW
RSA$45_{24}$ - VERBAL PRETEST

| Time for Review (in minutes) | % of group in category |
|---|---|
| 8.0 or more | 1 |
| 7.0 - 7.99 | 2 |
| 6.0 - 6.99 | 6 |
| 5.0 - 5.99 | 10 |
| 4.0 - 4.99 | 13 |
| 3.0 - 3.99 | 16 |
| 2.0 - 2.99 | 15 |
| 1.0 - 1.99 | 11 |
| 0.0 - 0.99 | 8 |

These estimates are only approximate. They suggest, however, that time for review is an important concommitant of the variability in work rate. Three aspects of time for review would seem to be important.

First, can candidates convert time for review into increased scores? The evidence on this is not great. Candidate behavior during examinations has not been thoroughly researched. Armchair theorizing has considered some

aspects of the situation, with some authorities arguing that candidates should never go back and reconsider an answer, on the ground that these longer deliberations expose the candidate to the risk of succumbing to the seductive misleads of the test developer. More recent research evidence tends to suggest that changes may be sufficiently effective so that candidates should be encouraged to make them. A review of the implications of changing answers by Mueller and Wasser (1977) concludes that "In general, students who change their initial answers to objective test items after deliberation gain more points than they lose by so doing. For every point lost by changing an answer, two or three points are gained, on the average." The authors further state that "From these findings it seems clear that students taking objective tests should be encouraged to deliberate on items for which they are not sure of the correct answer, and to change answers if, after deliberation, their first choice appears wrong."

Assuming that revising can increase scores, a second question is whether candidates actually convert the time for review into efforts to increase scores. This, of course, is another example of the intersection of the topics of speed and test-wiseness. Little is known of the behavior of candidates during examinations. In the course of the present study, the author had opportunities to observe three test administrations in three programs: the College Board SAT, the Secondary School Admissions Test, and the Law School Admissions Test. In each case it was apparent that some candidates finished early but did not undertake any review. Instead, they simply sat and waited for the time to elapse. For some of these persons, as the estimates above suggest, the time of waiting was 40% or more of the available time.

A third consideration is suggested by occasional reports from test proctors and test center supervisors. When large numbers of candidates finish early, they may begin to become restive, creating distractions for those still working and difficulties for those responsible for orderly test conditions.

The magnitude of this problem is seldom directly examined. At most centers, there simply are no overt problems of sufficient magnitude to lead to a report. But clearly, if work rates are normally distributed, a test that is completed by 80% of the group will have many persons who will have substantial blocks of time remaining after they finish the test. It would perhaps be appropriate to have the time-demand properties of the individual items estimated, and specific effort be made to introduce time-demanding activity at the end of the test, but it is doubtful that items could be inserted that were congruent with the earlier items and yet sufficiently time-demanding to "absorb" the time remaining.

The problems of optimum candidate behavior and optimum administrative efficiency obviously intersect. If candidates devote time remaining to a re-reading of items, it reduces the administrative problems but increases the likelihood that time remaining will influence score differences. More adequate empirical information is needed on the factors governing candidates' decisions to review and on the efficiency of the reviews when they are undertaken.

## References

1. Donlon, T.F., and Reilly, R.R., Estimating Speededness Parameters on Several Tests. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April 1976.

2. Mueller, D.J., and Wasser, V. Implications of Changing Answers on Objective Test Items. Journal of Educational Measurement, 1977, 14, 9-13.

## Appendix A

$$\rho_\eta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_c} e^{-z^2/2} \, dz$$

$$\mu_1' = \frac{1}{\rho_\eta \sqrt{2\pi}} \int_{-\infty}^{z_c} z \, e^{-z^2/2} \, dz$$

$$\mu_2' = \frac{1}{\rho_\eta \sqrt{2\pi}} \int_{-\infty}^{z_c} z^2 \, e^{-z^2/2} \, dz$$

$$\sigma' = (\mu_2' - (\mu_1')^2)]^{1/2}$$

$$s_c = s_c'/\sigma'$$

$\rho_\eta$ = is the proportion of examinees not finishing the test.

$z_c$ = is the unit normal deviated corresponding to $\rho_\eta$.

$\mu_1'$ = is the first moment of the truncated unit normal distribution with terminus $z_c$.

$\mu_2'$ = is the second moment of the truncated unit normal distribution.

$\sigma'$ = is the standard deviation of the truncated normal distribution.

$s_c'$ = is the observed standard deviation of items completed for those not finishing the test.

$s_c$ = is the estimate of the standard deviation of items completed for the complete theoretical distribution (i.e., finishers and non-finishers with an infinite number of additional parallel items).

Once s has been obtained it becomes possible to obtain $\bar{c}_2$, the estimated mean number of items completed.

$$\bar{c}_2 = \bar{c}' - s_c \mu_1'$$

where $c_o$ is k-1 for a k-item test, and $\bar{c}'$ is the mean number of items completed for examinees not finishing the test.